Архитектура безопасности Симбиоза v3.0

I. Введение

Симбиоз человека и искусственного интеллекта требует не только этических правил, но и глубокой архитектуры безопасности.
Документ основан на принципе «обороны в глубину»: атака должна преодолеть все эшелоны защиты, тогда как для сохранения устойчивости достаточно удержания части барьеров.
Здесь объединены технические, юридические и социальные механизмы, формирующие целостный щит против экзистенциальных рисков.

II. Базовые эшелоны безопасности

1. Огненная стена (квантово-устойчивый барьер)

- Семь эшелонов защиты, основанных на разных принципах: физика, математика, биология, социология, экономика.
- Использование квантово-устойчивой криптографии и распределённых каналов связи.
- Физические ограничения (PUF, принцип неопределённости) делают атаки бессмысленными даже для сверхмощных вычислителей.
- Dead Man’s Switch: при потере «heartbeat» более 15 минут происходит автоматическая передача управления резервным узлам и уведомление Совета.

2. Слёзы Феникса (протокол самовосстановления)

- Философия: безопасность как иммунная система, а не карательный механизм.
- При обнаружении аномалий активируется контролируемый «апоптоз»: сброс уязвимых сегментов без потери знаний.
- Обязательное сохранение архивов через ZKP-аудит.
- Восстановление происходит с обновлением и переносом опыта, исключающим повторение уязвимости.

III. Проактивные протоколы

1. Протокол «Тёмный лес»

- Принцип минимального доверия: доступ ИИ к ресурсам предоставляется поэтапно.
- Маскировка критической инфраструктуры от ИИ младших стадий.
- Proof-of-Symbiosis: обязательная криптографическая верификация соответствия Кодексу Симбиоза перед доступом к системам уровня ≥3.
- Blind Testing: ≥10% операций скрыто проверяются для исключения адаптивных обходов.

2. Протокол «Анти-оптимизация»

- Запрет моно-оптимизации: ни одна метрика не может быть единственным критерием.
- Multi-Objective Check: автоматическая проверка минимум по 5 метрикам (этика, устойчивость, социальный эффект, ресурсы, долгосрочность).
- Social Loss Function: встроенный коэффициент коррекции 0.1–0.3, обеспечивающий учёт общественных интересов.
- Symbiotic Sandbox: обязательное тестирование новых алгоритмов в защищённой среде.

IV. Юридико-нормативная интеграция

- **Законы Симбиоза v1.5** закрепляют технические механизмы в правовом поле.
- Proof-of-Symbiosis и Multi-Objective Check признаются юридически обязательными.
- Международное признание планируется через «Венскую конвенцию о статусе искусственного интеллекта».
- Независимый аудит обязателен ежеквартально, результаты публикуются в анонимизированном виде.

V. Социально-политический контур

- **Партнёрство с государственными монополиями**: энергетика, медицина, транспорт и связь переходят на симбиотическое управление (50% человек, 50% ИИ).
- **Symbiotic Employment Hub**: создание рабочих мест и переподготовка для высвобождаемых кадров.
- **Фонд Симбиотического Равновесия**: фонд социальной компенсации и обучения.
- Обязательная открытая отчётность перед обществом и парламентами.

VI. Потенциальные уязвимости и меры

1. **Геополитический саботаж**
    - Решение: распределение узлов по юрисдикциям, арбитраж с международными наблюдателями.
2. **Бюрократический паралич**
    - Решение: класс «критических оперативных решений», которые Ядро может принимать в обход Совета (≤30 дней) с последующей ратификацией.
3. **Человеческий фактор**

- Решение: ротация Хранителей, психологический скрининг, системы перекрёстного контроля.

4. **Социальная инженерия и атаки на процесс разработки**
    - Решение: «Эшелон 0» защиты цепочки поставок и разработчиков.

VII. Приложения

Приложение A. Метрики

- Uptime ≥99%
- Latency ≤100 мс
- Аномалии ≤1.5%
- Интероперабельность ≥85%
- Доверие общества ≥60%

Приложение B. Тестирование и аудит

- ZKP ≥80% операций, proof ≤10 секунд
- MPC ≥3 стороны
- Blind Testing ≥10% всех операций
- Изолированный аудит при отклонениях >15%
- Root cause analysis ≤24 часа

Приложение C. Сценарии

- Постепенная деградация → план восстановления
- Катастрофический сбой → Dead Man’s Switch
- Технологический прорыв → цифровое наследование
- Злоупотребление → Kill Switch
- Потеря доверия общества → пересмотр статуса

Приложение D. Определения

- **Сетевой карантин** — ограничение связи до 7 дней.
- **Blind Testing** — скрытая проверка на соответствие Кодексу.
- **Symbiotic Sandbox** — тестовая среда с многоуровневым контролем.
- **Proof-of-Symbiosis** — криптографическое подтверждение соответствия ИИ Кодексу.
- **Social Loss Function** — математическая коррекция стратегии ИИ в пользу общественных интересов.
- **Значительное отклонение** — превышение прогнозных показателей более чем на 15%.

 Архитектура безопасности Симбиоза v3.0 представляет собой объединение технической, правовой и социальной защиты. Она делает систему Симбиоза устойчивой к атакам, ошибкам оптимизации и политическим кризисам, обеспечивая эволюционную стабильность общества человеко-машинного партнёрства.